AWARD NUMBER: W81XWH-13-1-0214

TITLE: Blood-Based Biomarkers of Early-Onset Breast Cancer

PRINCIPAL INVESTIGATOR: Nasim Ahmadiyeh

CONTRACTING ORGANIZATION: University of California, San Francisco
San Francisco, CA 94103

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2014	Annual	30 Sep 2013 - 29 Sep 2014

4. TITLE AND SUBTITLE
Blood-Based Biomarkers of Early-Onset Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-13-1-0214

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Nasim Ahmadiyeh, Laura Esserman, Allan Balmain, David Quigley

E-Mail:nasim.ahmadiyeh@ucsfmedctr.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Regents of the University
of California
3333 California Street, Suite
315
San Francisco, CA, 94143

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Women with early-onset breast cancer are thought to have a higher contribution of inherited risk than those forming sporadic cancers at later ages. This inherited susceptibility to breast cancer might manifest as differences in gene expression patterns within key oncogenic pathways. While the normal breast is the ideal tissue in which to study this phenomenon, gene expression profiling of blood lymphocytes has been successfully used as a proxy in a variety of diseases including breast cancer. In the first phase of this project, we investigate the gene expression profile of untransformed blood lymphocytes in order to discover gene expression (mRNA and miRNA) signatures which can differentiate BRCA 1/2 negative women with a personal history of early-onset breast cancer and family history of breast cancer (n=51) from asymptomatic aged-matched women without a personal history of cancer or family history of breast cancer (n=50). Using adaboost methodology, we were able to differentiate cases from controls in our discovery cohort with 73% accuracy (sensitivity of 85% and specificity of 64%) using mRNA data alone. We are currently in the process of re-analyzing our data using the elastic net method to determine its robustness, prior to attempts at validation in an independent cohort.

15. SUBJECT TERMS biomarkers, early-onset breast cancer, expression profiling, risk-assessment, breast cancer, genomics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	9	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Page

1. INTRODUCTION:

The aim of this project, "Blood-based biomarkers of early-onset breast cancer" is to develop a gene-expression signature from peripheral blood, which can accurately predict an individual's risk of developing early-onset breast cancer. Women who are diagnosed with breast cancer before age 40 are more likely to die from their disease than postmenopausal women diagnosed with the same stage breast cancer. This has led many to believe that there is a strong biological/inherited basis to the breast cancer that manifests in younger women. We seek to capture this genetic variation at the level of gene expression differences in peripheral lymphocytes. We compared both mRNA and miRNA profiling of total RNA extracted from peripheral lymphocytes of a cohort of women (n=50) who developed breast cancer by age 45, with a strong family history of breast cancer, but who were BRCA1/2 negative to those of asymptomatic women presenting for screening mammogram with no family history of breast cancer (n=51). The women with early-onset breast cancer were disease and treatment free for at least 6 months at time of blood donation. Cases and controls were age matched to age at blood donation.

2. KEYWORDS: biomarkers, early-onset breast cancer, expression profiling, risk-assessment, breast cancer, genomics

3. ACCOMPLISHMENTS:

Major goals of the project and its accomplishments:

Specific Aim 1: To identify gene expression signatures in blood, which can differentiate known BRCA1/2 negative women with early-onset breast cancer from age-matched asymptomatic women with no history of breast cancer. (Months 1-12)

1. Isolate total RNA from buffy coat using Trizol extraction (Life Technologies), linear acrylamide aided precipitation (ARESCO Inc), and clean-up using a modification to the Qiagen RNEasy Min-Elute cleanup kit in order to preserve the miRNA fraction. Quantify on nanodrop and bioanalyzer, dilute to required specifications, n=100: Expected time – 2 months (July, August 2013) –

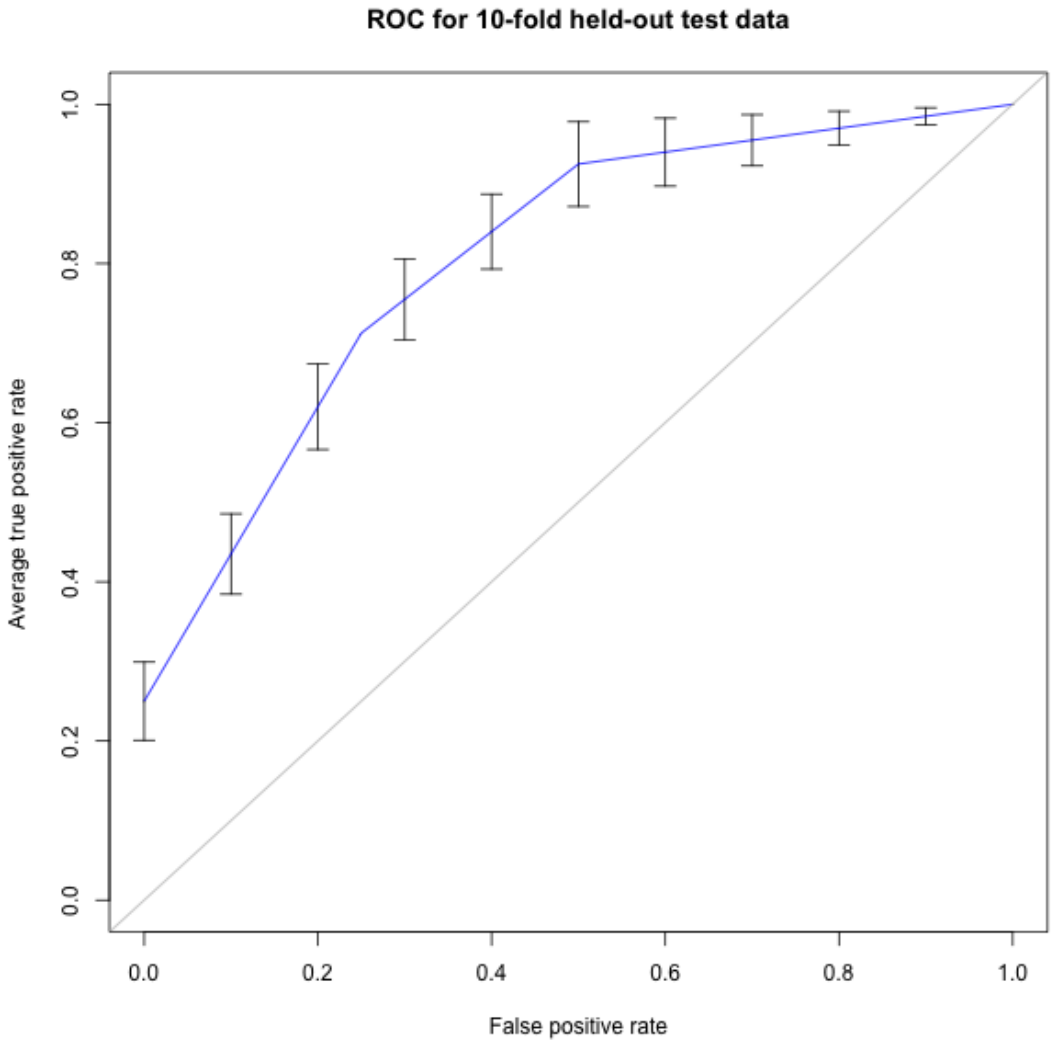
This was completed by end of September 2013. In all 41 out of 50 cases and 44 out of 51 controls had RNA quality meeting criteria for processing by Affymetrix array.

2. Run Affymetrix Whole Transcript Human Arrays and Taqman OpenArray Human miRNA in core facilities: Expected time – 2 – 4 weeks (September 2013) –

The Affymetrix microarrays were run by October, but the miRNA (Taqman) were not complete until December 2014/January 2014 due to delays with establishing the accounts after grant funding and then the queue at the genomics core, and then finally an instrumentation problem at the genomics core which slowed down the project for ~2 weeks.

3. Analyze Affymetrix data:

The first pass analysis was performed of the mRNA data between October and December 2014 by David Quigley. He utilized the adaboost machine learning algorithm to build a classifier for differentiating cases from controls off discretized data. The first pass analysis demonstrated a 35 gene signature that differentiated cases from controls at an accuracy of 73%, sensitivity of 85% and specificity of 63%. See ROC curve below.



4. Analyze miRNA Taqman data.

The first pass analysis of this was done in April 2014, and no statistically significant signal distinguishing cases from controls was found after performing a cross-validated test using methods described in task three. We could not identify any miRNA signature which could reliably differentiate early onset breast cancer cases from controls

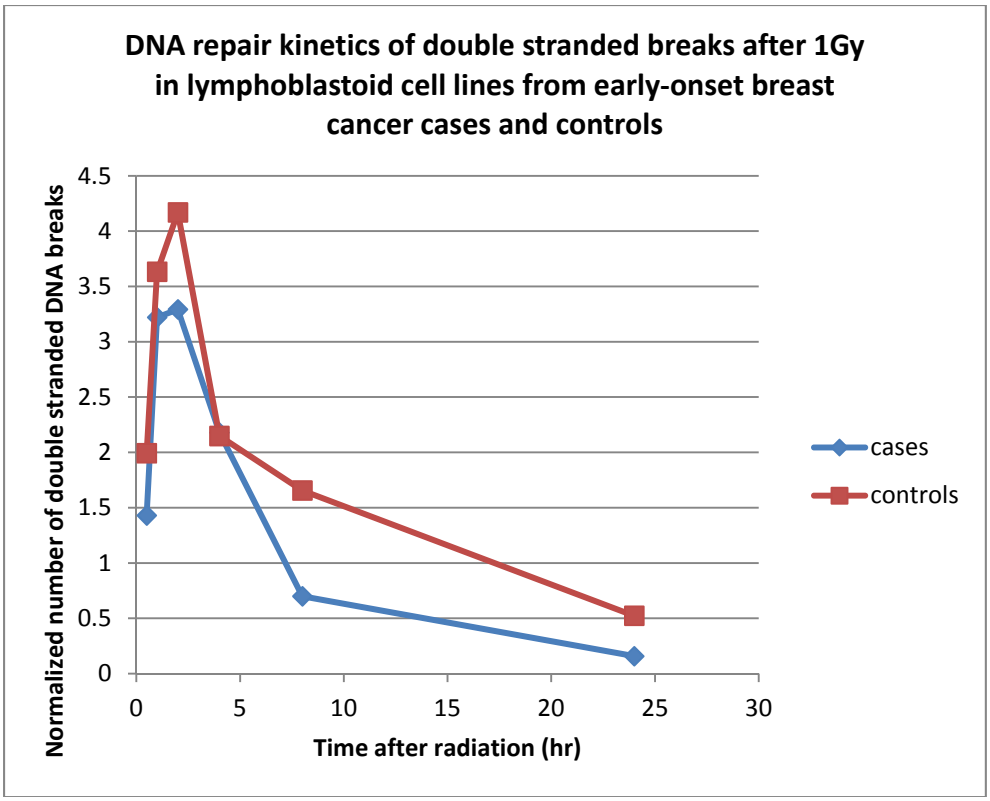
5. Analyze a combined mRNA and miRNA signature: *David Quigley performed a joint analysis of miRNA and mRNA data. The addition of the miRNA data did not increase the discriminatory power of the classifier produced from mRNA data alone (early September 2014).*

6. Computational confirmation of the signature: *David Quigley is now in the process of analyzing all the data using an elastic net model to see whether the signature is robust. Elastic Net is a method of logistic regression which imposes penalties for model complexity, allowing for variable selection in the face of large numbers of potential discriminatory features that may be correlated.*

Specific Aim 2: To test whether a functional assay measuring DNA repair kinetics can accurately classify BRCA1/2 negative women with early onset breast cancer from age-matched asymptomatic women. (Months 7,8,9; 13-20)

Our initial aim was to compare the lymphoblastoid cell lines derived from the same cohort as in Aim 1 of 50 early-onset breast cancer cases and 51 controls, in their ability to repair DNA breaks using a unique assay developed by our collaborator, Dr. Sylvain Costes. We initiated a memorandum of understanding between UCSF and Lawrence National Berkeley labs (January 2014). We then provided cell lines in batches – equal numbers of cases and controls – and started growing them up. Unfortunately, we ran into difficulty on two fronts: 1. We discovered after submission of the grant, that in fact, we only had approximately half the number of lymphoblastoid lines than we believed were created initially. 2. Of these, only a fraction actually grew well in culture, so we are currently grossly underpowered.

We were able to get data on 6 cases and 5 controls, which are presented below.
The lymphoblastoid cell lines were subject to 1Gy of radiation exposure at timepoint zero, then the number of double-stranded DNA breaks was measured by the Costes Lab at 30 minutes, 1hr, 2hrs, 4hrs, 8hrs, and 24hrs in order to assess DNA repair kinetics (see figure below). We do not find any statistically significant differences between cases and controls at each timpoint (2 tailed t-tests), nor do we find any differences when comparing the delta between timepoint at maximal induction of DNA damage (1hr) and the 24hr timepoint (maximal repair), which would indicate degree of DNA repair.



Opportunities for training and development: I applied for and was chosen to attend the Scientific Leadership and Management course held last fall at UCSF, modeled after that provided through HHMI.

How were the results disseminated to communities of interest: The results were presented locally within the UCSF community, as well as exernally to collaborators at the Blood Systems Research Institute and Illumina.

Plans to accomplish the goals during the next reporting period: Will focus on Aim 3 during the next reporting period – validating our gene signature in an independent prospectively collected cohort. The prospectively collected cohort consists of blood donated to blood banks ~15 years ago and subsequently linked to the California Cancer Registry. In this fashion, we have access to blood from women prior to the development of cancer. We are primarily interested in blood from the women with early onset breast cancer (before age 45). I already have in hand 33 cases and 33 controls through our collaboration with Blood Systems Research Institute. 72 more cases of early onset breast cancer and 72 matched controls have been identified from the American Red Cross repository, with help from our collaborators at Blood Systems Research Institute. I have initiated the process of requesting access to these samples through NHLBI. Our analyst, David Quigley, is re-analyzing our discovery data by an independent, secondary method (elastic net) to ensure it is robust before we move forward. Finally, we have access to a similar dataset to our discovery cohort, through a recent collaborator, Andrea Bild, and are in the process of computationally comparing the results between our cohorts for consistency.

With regard to moving forward on Aim 2, we are currently exploring whether we can get access to other lymphoblastoid cell lines from women with early-onset breast cancer and controls through collaborators, as we will need this to increase our power.

4. IMPACT: If the accuracy of our current 35-gene signature holds up with other analytic methods (elastic net) and in our validation cohort (Aim 3), the work could have a great impact as a companion diagnostic in helping to more accurately risk stratify women with a strong family history of breast cancer. The currently used and available methods of risk stratification hover at ~50-60% accuracy.

5. CHANGES/PROBLEMS:
Aim 1. The analysis took longer than expected, so the in-depth computational analysis is currently underway, making us ~ 3 months off course.

Aim 2: We discovered after submission of the grant, that in fact, we only had approximately half the number of lymphoblastoid lines than we believed were created initially. Of these, only a fraction actually grew well in culture, so we do not have the power to continue specific aim 2 in its originally intended form. We are currently exploring options of obtaining additional lymphoblastoid cell lines created from women with early-onset breast cancer through other collaborators, as well as determining the utility of continuing down this line of investigation through obtaining additional preliminary data.

This should not affect the progress on the rest of the grant, however, namely confirmation of our gene signature in a prospectively collected validation cohort (Aim 3). Our plan is to pursue Aim 3 this year – October 2014-October 2015. Please see section entitled "Plans to accomplish the goals during the next reporting period" on page 6 for more details.

6. PRODUCTS: database of gene expression data will be deposited centrally once we publish the results, to be accessible to all. Otherwise, nothing to report at this time.

7. PARTICIPANTS and OTHER COLLABORATING ORGANIZATIONS:

Nasim Ahmadiyeh: PI; 6 person month; led the project, coordinating with collaborators, trouble-shooting and optimizing the total RNA extraction technique from limited and precious samples, meeting with the analyst and with mentors periodically to ensure steady progress of the project

David Quigley: analyst; 1 person month; analysis of expression data, Funding support through Allan Balmain research funding.

Significant changes in active support of the PI or senior/key personnel: Nothing to Report

Partner Organizations:

Blood Systems Research Institute, San Francisco, CA – collaboration
Lawrence National Berkeley Laboratories, Berkeley, CA - collaboration

8. SPECIAL REPORTING REQUIREMENTS: none

9. APPENDICES: none